Acta Cybernetica, Vol. 11, No. 1-2, Szeged, 1993

On codes concerning bi-infinite words

Do Long Van* Nguyen Huong Lam* Phan Trung Huy*

Abstract

In this paper we consider a subclass of circular codes called Z-codes. Some tests of Sardinas-Patterson type for Z-codes are given when they are finite or regular languages. As consequences, we prove again the results of Beal and Restivo, relating regular Z-codes to circular codes and codes with finite synchronization delay. Also, we describe the structure of two-element Z-codes.

1 Preliminary

In this paper only very basic notions of free monoids and formal languages are needed. As a general reference we mention [7], and for the facts concerning codes we always refer to [3] silently. In addition to this we use also notions concerning infinite and bi-infinite words without very formal definitions because of a wide availability of papers on the subject. To fix our notations we want to specify the following. Throughout this paper A denotes a finite alphabet. The free monoid generated by A, or the set of finite words, is denoted by A^* and its neutral element, the empty word, by ε. As usual we set $A^+ = A^* - \varepsilon$. For a word x in A^*, $|x|$ means the length of x. We call a nonempty word x *primitive* if it is not a proper power of any word, otherwise x is *imprimitive*. We call two words x and y *copower* if they are powers of the same word. For example, as well known two different words are copower if and only if the set they form is not a code. For two finite words x and y the notation yx^{-1} and $x^{-1}y$ are used to denote the right and the left quotient of y by x respectively. Naturally, the quotient and the product of two words can be extended to languages, i.e. subsets of A^* :

$$\begin{aligned} X^{-1}Y &= \{x^{-1}y : x \in X, y \in Y\},\ YX^{-1} = \{yx^{-1} : x \in X, y \in Y\}. \\ XY &= \{xy : x \in X, y \in Y\}, X^2 = XX, \ldots; \end{aligned}$$

and $X^* = \bigcup_{n\geq 0} X^n$ (the Kleene closure of X).

In the following, our consideration is mainly based on the notion of infinite and bi-infinite words on A. Let ${}^{N}\!A$, A^N, A^Z be the sets of left infinite, right infinite and bi-infinite words on A respectively. For a language X of A^*, we denote ${}^{\omega}X$, X^{ω} and ${}^{\omega}X^{\omega}$ the left infinite, the right infinite and the bi-infinite product of nonempty words of X respectively, i.e. their elements are obtained by concatenation of words of $X-\varepsilon$ carried out infinitely to the left, to the right or infinitely in both directions. For example,

$${}^{\omega}X = \{\ldots u_2u_1 : u_i \in X - \varepsilon, i = 1, 2, \ldots\}.$$

*Institute of Mathematics, P.O.Box 631, 10000 Hanoi, Vietnam

Factorizations in elements of X (over X, on X) of a left or right infinite word are understood customarily (see [10] for details), but factorizations of a bi-infinite word need a special treatment as follows. Let $w \in A^Z$ be in the form:

$$w = \ldots a_{-2}a_{-1}a_0a_1a_2\ldots$$

with $a_i \in A$. A *factorization* on elements of X of the bi-infinite word w is a strictly increasing function $\mu : Z \longrightarrow Z$ satisfying $x_i = a_{\mu(i)+1}\cdots a_{\mu(i+1)} \in X$ for all $i \in Z$. Two factorizations μ and λ are said to be *equal,* denoted $\mu = \lambda$ if there is $t \in Z$ such that $\lambda(i+t) = \mu(i)$ for all $i \in Z$. Otherwise, λ and μ are *distinct,* denoted $\mu \neq \lambda$. It is easy to verify that $\mu \neq \lambda$ iff $\mu(Z) \neq \lambda(Z)$, or equivalently, there exist a word $u \in A^+$, two bi-infinite sequences of words of $X : \ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots$ and $\ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots$ such that

$$\begin{aligned} \ldots x_{-2}x_{-1}u &= \ldots y_{-1}y_0, \quad |u| \leq |x_0|, \\ x_0x_1\ldots &= uy_1y_2\ldots, \quad |u| \leq |y_0| \end{aligned}$$

with $u \neq x_0$ or $u \neq y_0$.

If every rigth infinite word of A^N has at most one factorization on elements of X then X is said to be an N-*code* (see [10], where in a wider context N-code is called *strict* code). Analogously, if every left infinite word possesses this property, we call X an $\overline{N}$-*code.* Obviously, X is an N-code iff $\overline{X} = \{\overline{x} : x \in X\}$ is an $\overline{N}$-code, where $\overline{x}$ is the mirror image of the word x. For the bi-infinite words, we have our basic

Definition 1 *A language X of A^+ is a Z-code if all factorizations on X of every bi-infinite word are equal.*

Example 1 Every singleton $\{u\}$ is always both an N-code and an $\overline{N}$-code but it is a Z-code if and only if u is primitive. The two-word language $X = \{ab, ba\}$ is both an N-code and an $\overline{N}$-code, but it is not a Z-code since the word ${}^{\omega}(ab)^{\omega}$ has two factorizations $\ldots ab.ab.ab\ldots$ and $\ldots ba.ba.ba\ldots$, which are verified directly to be distinct.

The family of Z-codes is closely connected with the so-called circular code [3]. A language X of A^* is said to be *circular* if for any $x_0, x_1, \ldots x_m, y_0, y_1, \ldots y_n$ of X and s, t of A^* the equalities

$$\begin{aligned} x_1x_2\ldots x_m &= ty_0\ldots y_m s, \\ x_0 &= st \end{aligned}$$

imply $s = \varepsilon, m = n$ and $x_0 = y_0, \ldots, x_m = y_m$.

It is easy to see that every circular language is a code and that every Z-code is a circular code. But not always a circular code is a Z-code, as the following code [4] $X = \{ab\} \cup \{ab^iab^{i+1}, i = 0, 1, 2, \ldots\}$ shows that. Nevertheless, every regular circular code is a Z-code i.e. the families of regular Z-codes and regular circular codes coincide, as shown by Beal [2]. Therefore, results and algorithms invented for circular codes can be applied to Z-codes. However, in the next section we work independently with Z-codes, proposing some tests for regular and finite Z-codes. As consequences of that, we can obtain a result of A. Restivo on codes with finite (bounded) synchronization delay [11] and the aforementioned Beal's result. Also, for completeness, as an easy consequence of [1], we describe the structure of two-word Z-codes.

2 Tests for Z-codes

We develop now a criterion to verify whether a finite subset X of A^+ is a Z-code. Our procedure is something like the Sardinas- Patterson one (cp. [10]), but actually instead of one sequence of subsets associated to X we need two sequences associated to each overlap of elements of X. Precisely, we define first the subset:

$$W(X) = \{w \in A^+ : \exists u, v \in A^*; \exists x, y \in X : uw = x, wv = y, uv \neq \varepsilon\}$$

whose element is called an *overlap* of elements of X. For each $w \in W(X)$, we define two sequences $U_i(w, X)$ an $V_i(w, X)$ of subsets of A^* as follows

$$\begin{aligned} U_0(w,X) &= w^{-1}X - \{\varepsilon\}, \\ U_{i+1}(w,X) &= U_i(w,X)^{-1}X \cup X^{-1}U_i(w,X), \\ V_0(w,X) &= Xw^{-1} - \{\varepsilon\}, \\ V_{i+1}(w,X) &= XV_i(w,X)^{-1} \cup V_i(w,X)X^{-1}, \end{aligned}$$

$i = 0, 1, 2, \ldots$. Further, if there is no risk of confusion, instead of $W(X)$, $U_i(w,X)$, $V_j(w,X)$ we write simply W, U_i, V_j. The following property of $U_i(w,X)$, $V_j(w,X)$ is useful in the sequel.

Lemma 1 *For every $N \geq 0$ and for any word u, $u \in U_N(w,X)$ iff there exist $x_1, \ldots, x_n, y_1, \ldots, y_m \in X$ such that $m + n - 1 = N$ and either*

$$wx_1 \ldots x_n = y_1 \ldots y_m u, \quad |u| \leq |x_n|, |w| < |y_1|$$

or

$$wx_1 \ldots x_n u = y_1 \ldots y_m, \quad |u| \leq |y_m|, |w| < |y_1|.$$

Remark. Similarly, the symmetrical statement holds for V_j.

Proof. By induction on N. For $N = 0$ we have

$$u \in U_0 \Leftrightarrow (\exists y_1 \in X : w^{-1}y_1 = u \Leftrightarrow wu = y_1, |u| < |y_1|, |w| < |y_1|).$$

Suppose the lemma is true for some $N \geq 0$, we prove it true for $N + 1$. We have

$$u \in U_{N+1} \Leftrightarrow \exists u' \in U_N, \exists x \in X : u'u = x \vee xu = u'.$$

By induction hypothesis, $u' \in U_N$ iff there exist $x_1, \ldots, x_n, y_1, \ldots, y_m \in X$ such that $n + m - 1 = N$ and either

$$wx_1 \ldots x_n u' = y_1 \ldots y_m, \quad |u'| \leq |y_m|, \quad |w| < |y_1| \tag{1}$$

or

$$wx_1 \ldots x_n = y_1 \ldots y_m u', \quad |u'| \leq |x_n|, \quad |w| < |y_1|. \tag{2}$$

Therefore $u \in U_{N+1}$ is equivalent to the fact that there exist $x_1, \ldots, x_n$, x, $y_1, \ldots, y_m$ in X such that

$$((u'u = x)\&((1) \vee (2))) \vee ((xu = u')\&((1) \vee (2)))$$

or equivalently

$$((u'u = x)\&(1)) \vee (u'u = x)\&(2))\vee$$
$$((xu = u')\&(1)) \vee ((xu = u')\&(2)).$$

The last, in its turn, as it is easy to verify, is equivalent to the fact that there exist $x_1, \dots, x_{n'}, y_1, \dots, y_{m'}$ in X such that $n' + m' - 1 = N + 1$ and

$$wx_1 \dots x_{n'} = y_1 \dots y_{m'}u, \quad |u| \leq |x_{n'}|, \quad |w| < |y_1|$$

or

$$wx_1 \dots x_{n'}u = y_1 \dots y_{m'}, \quad |u| \leq |x_{m'}|, \quad |w| < |y_1|,$$

i.e. the lemma is true also for $N + 1$.

Now we state a sufficient condition for a language to be a Z-code.

Proposition 1 *A finite subset X of A^+ is a Z-code if for every overlap w of elements of X, the following conditions hold:*

(i) if $w \in W \cap X$ then $U_i = \emptyset$ and $V_j = \emptyset$ for some $i, j \geq 0$;

(ii) if $w \in W - X$ then $U_i = \emptyset$ or $V_j = \emptyset$ for some $i, j \geq 0$.

Proof. We suppose that X is not a Z-code, i.e. at least one word of A^Z possesses two distinct factorizations on X, therefore we have two equalities:

$$\dots x_{-2}x_{-1}w = \dots y_{-1}y_0 \tag{1}$$

$$x_0x_1 \dots = wy_1y_2 \dots \tag{2}$$

for some $w \in A^+, |w| \leq |y_0|$ and $|w| \leq |x_0|, w \neq x_0$ or $w \neq y_0$, hence $w \in W$.

If $w \in W \cap X$ and, say, $w \neq x_0$, then $U_0 \neq \emptyset$. By (2), for every $N > 0$ there is the least integer $n \geq 0$ such that $|x_0 \dots x_n| \geq |wy_1 \dots y_N|$, that is

$$x_0x_1 \dots x_n = wy_1 \dots y_N u$$

for some word $u \in A^*, |u| < |x_n|$. By Lemma 1, $u \in U_{N+n}$. Thus $U_N \neq \emptyset$: (i) does not hold. For the case $w \neq y_0$, by (1) and the symmetrical version of Lemma 1 we get $V_N \neq \emptyset$ for all $N \geq 0$: (i) does not hold again.

Now let $w \in W - X$ then we have both $w \neq x_0$ and $w \neq y_0$. By the same argument as above we obtain $U_i \neq \emptyset$ and $V_j \neq \emptyset$ for all $i, j \geq 0$: (ii) does not hold. The proof is completed.

In order to make a converse of Proposition 1 for finite languages we prove a lemma, which places an upperbound on the least i such that $U_i = \emptyset$. For a finite subset $X = \{x_1, x_2, \dots, x_n\}$ of A^* we define $\| X \| = \sum_{i=1}^n |x_i|$. Note that each U_i consists only of right factors (i.e. suffices) of words in X and if $U_k = U_l \neq \emptyset$ for $k \neq l$ then $U_i \neq \emptyset$ for all $i \geq 0$. Since the set of right factors of words in X is of cardinality at most $\| X \|$, such an upperbound obviously exists and we can take it as $2^{\|X\|}$. In the following lemma a more refined estimation is given.

Lemma 2 *For any finite subset X of A^* and $w \in W$, the following assertions are equivalent*

(i) $U_i(w, X) \neq \emptyset$ for some $i \geq \| X \|$;

(ii) $U_i(w, X) \neq \emptyset$ for all $i \geq 0$;

(iii) There exist infinite sequences $x_1, x_2, \dots; y_1, y_2, \dots$ of words in X such that

$$wx_1x_2 \dots = y_1y_2 \dots$$

with $|w| < |y_1|$.

Remark. The symmetrical statement holds for $V_j(w,X)$.

Proof. (iii) $\Rightarrow$ (ii): already done in the proof of Proposition 1.

(ii) $\Rightarrow$ (i): obvious.

(i) $\Rightarrow$ (iii): Let $u_N \in U_N(w,X), N \geq \| X \|$. Then there exist $u_i \in U_i(w,X)$ such that $u_0 = w, u_{i+1} \in u_i^{-1}X$ or $X^{-1}u_i, i = 0,1,\ldots,N-1$. It is easy to see that $u_0, u_1, \ldots, u_N$ are suffices of words in X and the cardinality of the set of the suffices of the finite set X does not exceed $\| X \|$ and thus is less than $N+1$. Therefore, there are p and $q, 0 \leq p < q \leq N$ such that $u_p = u_q$. Let l be the largest number not exceeding $q-p$ such that $u_{p+1} = y_1^{-1}u_p, u_{p+2} = (y_1y_2)^{-1}u_p, \ldots, u_{p+l} = (y_1\ldots y_l)^{-1}u_p$, where $y_1,\ldots,y_l \in X$; otherwise $l = 0$. Then $u_{p+l+1} \in u_{p+l}^{-1}X$ and we apply Lemma 1 to the case $u_q \in U_{q-p-l}(u_{p+l},X)$ to obtain some words $x_1,\ldots,x_n$ and $z_1,\ldots,z_m$ of X such that

$$u_{p+l}x_1\ldots x_n = z_1\ldots z_m u_q$$

or

$$u_{p+l}x_1\ldots x_n u_q = z_1\ldots z_m.$$

Whence

$$u_p x_1\ldots x_n = y_1\ldots y_l z_1\ldots z_m u_q$$

or

$$u_p x_1\ldots x_n u_q = y_1\ldots y_l z_1\ldots z_m.$$

Since $u_p = u_q$, these equalities lead respectively to the infinite words

$$u_p(x_1\ldots x_n)^\omega = (y_1\ldots y_l z_1\ldots z_m)^\omega \tag{1}$$

or

$$u_p(x_1\ldots x_n y_1\ldots y_l z_1\ldots z_m)^\omega = (y_1\ldots y_l z_1\ldots z_m x_1\ldots x_n)^\omega. \tag{2}$$

On the other hand, since $u_p \in U_p(w,X)$, again by Lemma 1 we have

$$wx = y'yu_p, \quad |w| < |y'| \tag{3}$$

or

$$wxu_p = y'y, \quad |w| < |y'| \tag{4}$$

where $y' \in X, x, y \in X^*$. Combining (3) and (4) with (1) and (2), we get four possibilities that all lead to the desired infinite equality in (iii). Lemma 2 is proved.

Now we are ready to state our criterion.

Theorem 1 *A finite subset x of A^+ is a Z-code if and only if for every overlap w of elements of X, the following conditions hold:*

(i) if $w \in W \cap X$ then $U_i(w,X) = \emptyset$ and $V_j(w,X) = \emptyset$ for some $i, j < \| X \|$;

(ii) if $w \in W - X$ then $U_i(w,X) = \emptyset$ or $V_j(w,X) = \emptyset$ for some $i, j < \| X \|$.

Proof. The sufficient part is Proposition 1, we have to prove only the necessary one. Suppose that (i) or (ii) does not hold. We shall derive from this two equalities which show that X is not a Z-code. In fact, by Lemma 2 and its symmetrical version, we have two cases: there exist

(1) $w \in W \cap X$ and $x_i, y_j \in X, i, j = 0, 1, 2, \dots$ such that

$$x_0 x_1 \cdots = w y_0 y_1 \dots, \quad |w| < |x_0|$$

or

$$\dots x_1 x_0 = \dots y_1 y_0 w, \quad |w| < |x_0|;$$

(2) $w \in W - X$ and $x_i, y_j \in X, i, j = \cdots -2, -1, 0, 1, 2, \dots$ such that

$$x_0 x_1 \cdots = w y_0 y_1 \dots, \quad |w| < |x_0|$$

and

$$\dots x_{-1} x_0 = \dots y_{-1} y_0 w, \quad |w| < |x_0|$$

regarding (i) or (ii) does not hold.

The first case together with the obvious equalities $\dots ww = \dots ww$ and $ww \dots = ww \dots$ show that X is not a Z-code.

The equalities in the second case themselves ensure that X is not a Z-code. The proof is completed.

We give now some examples illustrating the execution of the algorithm.

Example 2 (a) Consider $X = \{a^2 b, b^2 a\}$. We apply Theorem 1 to show that X is a Z-code.

$$\begin{aligned} W &= \{a, b\}, \\ U_0(a, X) &= \{ab\}, \ U_1(a, X) = \emptyset, \\ U_0(b, X) &= \{ba\}, \ U_1(b, X) = \emptyset. \end{aligned}$$

Since $a, b \notin X$, we conclude that X is a Z-code.

(b) Let $X = \{u\}$ with u imprimitive, $u = \lambda^n (n \geq 2)$. Clearly $\lambda \in W - X$, $U_0(\lambda, X) = \{\lambda^{n-1}\}$, which implies $\lambda \in U_1(\lambda, X), \lambda^{n-1} \in U_2(\lambda, X), \dots$. Thus $U_i(\lambda, X) \neq \emptyset$ for all $i \geq 0$. So $\{u\}$ is not a Z-code.

Conversely, let $X = \{u\}$ not be a Z-code and let λ be an overlap of X such that $U_i(\lambda, X) \neq \emptyset$ for all $i \neq 0$. Since λ is an overlap of u, we have $x\lambda = u$ for some $x \in A^+$. Further, if $\lambda_0 \in U_0(\lambda, X)$ then $\lambda\lambda_0 = u$. Hence $U_0(\lambda, X) = \{\lambda_0\}$. Let $\lambda_1 \in U_1(\lambda, X)$ then $\lambda_0 \lambda_1 = u$. Thus $|\lambda_1| = |\lambda|$ and from $x\lambda = u$ it follows $\lambda = \lambda_1$. Consequently $\lambda_0 \lambda = \lambda \lambda_0 = u$, which with $\lambda_0, \lambda_1 \neq \varepsilon$ yield that u is imprimitive. Thus $\{u\}$ is a z-code if and only if u is primitive.

The main setback of Theorem 1 is that it is unfit for infinite (even regular) languages.

Example 3 Consider $X = \{a, cab, c, bc^+ d\}$ on the alphabet $A = \{a, b, c, d\}$. It is an infinite regular Z-code, but for all $i \geq 0 : U_i(c, X) \neq \emptyset$.

Nevertheless, for the important class of regular languages we can work out another algorithm close to the previous one, also of Sardinas-Patterson type. Let

X be a regular language and as before W be the set of overlaps. First, for each overlap $w \in W$ we construct two sequences:

$$\overline{U}_0 = w^{-1}X - \{\varepsilon\}, \qquad \overline{U}_{i+1} = \overline{U}_i^{-1}X^*,$$
$$\overline{V}_0 = Xw^{-1} - \{\varepsilon\}, \qquad \overline{V}_{i+1} = X^*\overline{V}_i^{-1}$$

for all $i \geq 0$, which, if needed, will be referred to as $\overline{U}_i(w, X)$ and $\overline{V}_j(w, X)$. Of course there is no need to compute $\overline{U}_i(w, X), \overline{V}_j(w, X)$ for all $w \in W$, it is sufficient to take representatives modulo the right and left principal congruence defined by X^* or X. Recall that for a subset X of A^* the following equivalence relation

$$u \equiv_R v \Leftrightarrow u^{-1}X = v^{-1}X, \quad u, v \in A^*,$$

called *right principal congruence* defined by X. Analogously is defined the *left principal congruence* $\equiv_L$. When X is regular, the number of right (left) principal congruence classes, called *right index* (resp. *left index*) of X, is finite and equal to the number of states of the minimal automaton recognizing X. Now we state

Theorem 2 *Let X be a regular subset of A^+ and m, e be the right and left index of X^*. Then X is a Z-code if and only if for all $w \in W$ the following conditions hold*

(i) $w \in W \cap X$ *implies* $\overline{U}_i(w, X) = \emptyset$ *and* $\overline{V}_j(w, X) = \emptyset$ *for some* $i < m, j < e$;

(ii) $w \in W - X$ *implies* $\overline{U}_i(w, X) = \emptyset$ *or* $\overline{V}_j(w, X) = \emptyset$ *for some* $i < m, j < e$.

Remark. As seen from the proof below, (i) and (ii) are sufficient for any language of A^* to be a Z-code.

Proof. In fact, we prove an equivalent statement: X is not a Z-code iff (i) or (ii) does not hold.

First, let X not be a Z-code. Then there exist two equalities:

$$\ldots x_{-2}x_{-1}w = \ldots y_{-1}y_0, \tag{1}$$

$$x_0x_1 \ldots = wy_1y_2 \ldots \tag{2}$$

with $|w| \leq |x_0|, |w| \leq |y_0|, x_i, y_j \in X, w \neq x_0$ or $w \neq y_0$, hence $w \in W$.

If $w \in W \cap X$, we assume for certainty that $w \neq y_0$ and consider (1), putting $v_0 = y_0w^{-1} \in \overline{V}_0$. From (1) we get

$$\ldots x_{-2}x_{-1} = \ldots y_{-2}y_{-1}v_0.$$

Choose $n \in N$ such that $|x_{-n} \ldots x_{-2}x_{-1}| \geq |v_0|$ and put again $v_1 = (x_{-1} \ldots x_{-1})v_0^{-1}$, hence $v_1 \in X^*v_0^{-1} \subseteq X^*\overline{V}_0^{-1} = \overline{V}_1$ and

$$\ldots x_{-(n+2)}x_{-(n+1)}v_1 = \ldots y_{-2}y_{-1}.$$

We apply this argument over and over again to see that $\overline{V}_j \neq \emptyset$ for all $j \geq 0$, i.e (i) does not hold.

If now $w \in W - X$, we have both $w \neq x_0$ and $w \neq y_0$. Similarly, we apply the argument above to (1) and (2) to verify $\overline{U}_i \neq \emptyset$ and $\overline{V}_j \neq \emptyset$ for all $i, j \geq 0$: (ii) does not hold.

Conversely, let $\overline{U}_i \neq \emptyset$ for all $i \geq 0$ and N be any integer not less than m, and $u_N \in \overline{U}_N$. There exist $u_i \in \overline{U}_i, i = 0, 1, \ldots, N-1$ such that $u_0 \in w^{-1}X$, $u_{i+1} \in u_i^{-1}X^*, i = 0, 1, \ldots, N-1$, or equivalently, $wu_0 \in X, u_iu_{i+1} \in X^*$, $i = 0, 1, \ldots, N-1$. Among $u_0, u_1, \ldots, u_N$ we can pick out u_q and u_p such that $p < q$ and $u_q \equiv_R u_p \bmod X^*$. We define now an infinite sequence of words $u'_0, u'_1, \ldots$ by putting

$$u'_i = u_i, \quad 0 \leq i \leq q-1$$

and

$$u'_{q+i} = u_{p+t}, \quad i = 0, 1, \ldots,$$

where t is the least nonnegative residue of i mod $q-p$.

It is easy to verify that

$$x'_i = u'_iu'_{i+1} \in X^*$$

for $i = 0, 1, 2, \ldots$ and

$$x' = wu'_0 = wu_0 \in X.$$

Consider now the infinite product $wu'_0u'_1 \ldots$ written in two ways

$$(wu'_0)(u'_1u'_2)\cdots = w(u'_0u'_1)(u'_2u'_3)\ldots$$

or

$$x_0x_1 \cdots = wy_1y_2 \ldots \tag{3}$$

with $x_0 \in X, |w| < |x_0|; x_i, y_j \in X^*$.

Analogously, if $\overline{V}_j \neq \emptyset$ for all $i \geq 0$, we have the equality

$$\ldots x_{-2}x_{-1}w = \ldots y_{-1}y_0, \tag{4}$$

where $y_0 \in X, |w| < |y_0|; x_i, y_j \in X^*$.

If now $w \in W \cap X$ and (i) does not hold, for instance, $\overline{U}_i \neq \emptyset$ for all i. Then (3) together with the obvious equality $\ldots ww = \ldots ww$ show that X is not a Z-code.

If $w \in W - X$ and (ii) does not hold, i.e. $\overline{U}_i, \overline{V}_j \neq \emptyset$ for all $i, j \geq 0$. Then (3) and (4) will give rise to two distinct factorizations on X of some bi-infinite word: X is not a Z-code and the theorem follows.

Example 4 We use Theorem 2 to show that the language $X = \{a, cab, c, bc^+d\}$ given in Example 3 is in fact a Z-code.

$$\begin{aligned}
W &= \{c, b\}, \\
\overline{U}_0(c, X) &= \{ab\}, \overline{U}_1(c, X) = c^+dX^*, \overline{U}_2(c, X) = \emptyset, \\
\overline{V}_0(c, X) &= \emptyset, \\
\overline{U}_0(b, X) &= c^+d, \overline{U}_1(b, X) = \emptyset.
\end{aligned}$$

Since $c \in W \cap X, b \in W - X, X$ is a Z-code.

In general Theorem 2 is not true for arbitrary languages, as shown in the following

Example 5 Consider $X = \{a^{i+2}ba^ib : i = 0,1,2,\ldots\} \cup \{ba^{2i+1}b : i = 0,1,2,\ldots\} \subseteq \{a,b\}^*$. Clearly, b is an overlap and for all $i \geq 0$, we have $ab \in \overline{U}_i(b,X)$, $a^{2(i+1)}b \in \overline{V}_i(b,X)$, i.e. $\overline{U}_i, \overline{V}_j \neq \emptyset$ for all $i,j \geq 0$, but a simple verification ensures that X is a Z-code.

We should mention two other algorithms to verify whether a regular code X is a Z-code. Both of them consist in checking the emptiness problem for some automata (Devolder and Timmerman [4], Beal [2]) that has as well known a polynomial time complexity in the number of states of automata.

Using Theorem 2 we give alternative proofs of the results of M.P. Beal and A. Restivo. First, we prove

Corollary 1 (M.P. Beal [1]) *Let X be a regular code. Then X is a Z-code if and only if it is a circular code.*

Proof. First, observe that if X is a code then

(1) for any $w \in W \cap X$: $\overline{U}_i(w,X) \cap X^* = \emptyset$ and $\overline{V}_i(w,X) \cap X^* = \emptyset$ for all $i = 0,1,2,\ldots$;

(2) for any $w \in W - X$: $\overline{U}_i(w,X) \cap X^* = \emptyset$ or $\overline{V}_i(w,X) \cap X^* = \emptyset$ for all $i = 0,1,2,\ldots$

that are trivially to be verified using Lemma 1 or its symmetrical version.

Let now X be a regular circular code, hence a code: (1) and (2) are satisfied.

Suppose that for some $w \in W \cap X$ we have, say, $\overline{U}_i \neq \emptyset$ for all $i = 0,1,2,\ldots$. For any $N \geq 0$ there exist $u_0, u_1, \ldots, u_{N-1}, u_N$ such that $u_1 \in u_0^{-1}X^*, \ldots, u_N \in u_{N-1}^{-1}X^*$. Since X^* is of finite right index m, if we take N sufficiently large, we can find $i,j : 0 \leq i < j$, such that $u_i^{-1}X^* = u_j^{-1}X^*$ and $j-1$ is even. Consider the words

$$u = u_{i+1} \ldots u_j, \quad v = u_{i+2} \ldots u_{j-1},$$

it follows $u_j u_{i+1} \in X^*, v \in X^*$ and $u = u_{i+1} v u_j \in X^*$. By circularity of X we get $u_j, u_{i+1} \in X^*$, in particular, $u_j \in \overline{U}_j \cap X^* \neq \emptyset$ contradicting (1). Therefore for any $w \in W \cap X$ we have $\overline{U}_i = \emptyset$ for some i and analogously $\overline{V}_j = \emptyset$ for some j.

As for any $w \in W - X$, by the same way, we can conclude that either $\overline{U}_i = \emptyset$ for some i or $\overline{V}_j = \emptyset$ for some j.

By virtue of Theorem 2, X is a Z-code. The proof is completed.

We now deduce another statement concerning codes with bounded synchronization delay. Recall that a subset X of A^* is said to be a *code with bounded synchronization delay* provided it is a code and for some integer $p \geq 0$, for all $u, v \in X^p$, and for all $g, f \in A^*$,

$$gu, vf \in X^*$$

whenever

$$guvf \in X^*.$$

The least number p satisfying this condition is the *synchronization delay* of X. The fact that every code with bounded synchronization delay is a Z-code is obvious, but the reverse conclusion is not always valid. A lot of interesting properties of these codes have been discovered, for example, in the finite case, these codes are exactly the very pure codes, i.e. circular codes (see [11], [12]). We have the following

Corollary 2 (A. Restivo [11]) *Let X be a regular subset of A^+, X is a code with bounded delay if and only if it is a Z-code satisfying $A^*X^dA^* \cap X = \emptyset$ for some positive integer d.*

Proof. "Only if" part: first, the fact that each code with bounded synchronization delay is a Z-code is easy. Further, we show that $A^*X^dA^* \cap X = \emptyset$ for all d exceeding the right index of X. Suppose on the contrary that

$$ux_1 \dots x_d v \in A^*X^dA^* \cap X$$

for some $x_1, x_2, \dots, x_d \in X$ and $u, v \in A^*$. Then, indeed, there exist i and $j, i < j \le d$, such that $ux_1 \dots x_i \equiv_R ux_1 \dots x_j \bmod X$ which implies that for all $k = 0, 1, 2, \dots$:

$$ux_1 \dots x_i(x_{i+1} \dots x_j)^k \equiv_R ux_1 \dots x_i(x_{i+1} \dots x_j)^{k+1} \bmod X$$

and consequently

$$ux_1 \dots x_i(x_{i+1} \dots x_j)^k x_{j+1} \dots x_d v \in X.$$

Hence the synchronization delay of X cannot be bounded.

Conversely, let X be a regular Z-code and $A^*X^dA^* \cap X = \emptyset$ for some positive integer d, hence $d \ge 2$. By Theorem 2, for all overlaps $w \in W, \overline{U}_m(w, X) = \emptyset$ or $\overline{V}_e(w, X) = \emptyset$, where m and e are the right and left index of X^*, respectively. We show that X is of bounded synchronization delay not greater than $p = (m+1)d$ (the value in [11] is $2(m+1)d$). If that is not so, there must exist some words $g, h \in A^*, x_1, \dots, x_p, x_{p+1}, \dots x_{2p}, y_1, \dots, y_q \in X$ such that

$$gx_1 \dots x_p x_{p+1} \dots x_{2p} h = y_1 \dots y_q \tag{1}$$

and for all $k = 1, 2, \dots, q$

$$gx_1 \dots x_p \ne y_1 \dots y_k.$$

Thus, it has to exist a unique positive integer $l \le q$ such that

$$y_1 \dots y_{l-1} < gx_1 \dots x_p < y_1 \dots y_l$$

and the largest positive integer $i \le p-1$ and the smallest positive integer $j \ge p+1$ satisfying

$$gx_1 \dots x_i \le y_1 \dots y_{l-1} < gx_1 \dots x_p < y_1 \dots y_l \le gx_1 \dots x_j \tag{2}$$

(abusing language, we write for words $x, y, x \le y, x < y$ to indicate that x is a prefix, a proper prefix of y, respectively). Since $y_l \notin A^*X^dA^*, j \le d+p$ and $i \ge p-d$.

Further, if in (2) $gx_1 \dots x_i = y_1 \dots y_{l-1}$ and $gx_1 \dots x_j = y_1 \dots y_l$ then

$$y_l = x_{i+1} \dots x_j, \quad j - i \ge 2$$

that is a contradiction with the fact that X is a code.

Alternatively, assume that $gx_1 \dots x_j \ne y_1 \dots y_l$ which gives rise to

$$gx_1 \dots x_{j-1}w = y_1 \dots y_l, \tag{3.1}$$

$$x_j x_{j+1} \dots x_{2p}h = wy_{l+1} \dots y_q, \tag{3.2}$$

where $w \in W$ and $|w| < |y_l|, |w| < |x_j|$. Similarly, the case $gx_1 \dots x_i \neq y_1 \dots y_{l-1}$ gives rise to

$$gx_1 \dots x_{i+1} = y_1 \dots y_{l-1}w, \tag{4.1}$$

$$wx_{i+2} \dots x_{2p}h = y_l y_{l+1} \dots y_q, \tag{4.2}$$

where $w \in W$ and $|w| < |y_l|, |w| < |x_{i+1}|$.

We will show that (3.2) or (4.2) equally leads to $\overline{U}_{2m}(w, X) \neq \emptyset$ and (3.1) or (4.1) – to $\overline{V}_k(w, X) \neq \emptyset$ with k abitrarily large, in particular $k \geq e$ that is quite a contradiction.

First, suppose that we have (3.2), setting

$$u_1 = x_{j+1} \dots x_{j+d}, \dots, u_m = x_{j+(m-1)d+1} \dots x_{j+md}$$

and let $q(k)$ the smallest integer such that for $k = 0, 1, \dots, m$

$$x_j u_1 \dots u_k \leq wy_{l+1} \dots y_{q(k)} \tag{5}$$

(for compactness, we set by convention that $x_j u_1 \dots u_k = x_j$ when $k = 0$). Since $A^* X^d A^* \cap X = \emptyset$, $w < x_j$ and $u_1, \dots, u_m \in X^d$, it follows $l+1 \leq q(0) < q(1) < \dots < q(m)$. Putting $v_k = y_{l+1} \dots y_{q(k)}, k = 0, 1, 2, \dots, m$, by (5) and $A^* X^d A^* \cap X = \emptyset$ we get

$$x_j u_1 \dots u_k \leq wv_k < x_j u_1 \dots u_{k+1} \tag{6}$$

for $k = 1, 2, \dots, m-1$ and

$$wv_{k-1} \leq x_j u_1 \dots u_k \leq wv_k \tag{7}$$

for $k = 1, \dots, m$.

It is easy to verify that (6), (7) together with $w < x_j$ yield

$$(wv_0)^{-1}(x_j u_1) \in \overline{U}_2, \dots, (wv_{m-1})^{-1}(x_j u_1 \dots u_m) \in \overline{U}_{2m},$$

i.e. $\overline{U}_{2m} \neq \emptyset$.

Likewise, since $i + 2 \leq j$, (4.2) leads also to $\overline{U}_{2m} \neq \emptyset$.

Now, as far as $\overline{V}_e$ is concerned, we treat (3.1) and (4.1) as above, only in the symmetrical way. Directly, (3.1) or (4.1) cannot lead to $\overline{V}_e \neq \emptyset$, but we can "pump" them up to some equalities "long" enough by proceeding as follows. Suppose, for example, that we have (4.1). Among $m + 1$ numbers $1, d + 1, \dots, md + 1$ there must exist a, b such that $gx_1 \dots x_a \equiv_R gx_1 \dots x_b \bmod X^*$ with $a < b$. Note that $b - a \geq d \geq 2$ and $a, b \leq md + 1 \leq p - d + 1 \leq i + 1$. Further, for some integer $s \leq t \leq l$ we must have

$$\begin{aligned} y_1 \dots y_{s-1}u_a &= gx_1 \dots x_a, \\ gx_1 \dots x_a v_a &= y_1 \dots y_s, \\ u_a v_a &= y_s \end{aligned}$$

and

$$\begin{aligned} y_1 \dots y_{t-1} u_b &= g x_1 \dots x_b, \\ g x_1 \dots x_b v_b &= y_1 \dots y_t, \\ u_b v_b &= y_t, \end{aligned}$$

where $u_a, v_a, u_b, v_b \in A^*$. Hence $x_{a+1} \dots x_{i+1} \in v_a X^* w$. From $g x_1 \dots x_a \equiv_R g x_1 \dots x_b \bmod X^*$ it follows

$$g x_1 \dots x_a \equiv_R g x_1 \dots x_a (x_{a+1} \dots x_b)^k \bmod X^*$$

for all $k = 0, 1, 2, \dots$. Since $g x_1 \dots x_a v_a \in X^*$ we have $g x_1 \dots x_a (x_{a+1} \dots x_b)^k v_a \in X^*$. Therefore

$$g x_1 \dots x_a (x_{a+1} \dots x_b)^k x_{a+1} \dots x_{i+1} \in X^* w, \tag{8}$$

where, as before, $|w| < |x_{i+1}|$.

Looking into (8) we see that the left-hand side of (4.1) is pumped up by a product of $k(b - a - 1)$ words. We take k large enough to obtain a sufficiently "long" equality of the form (4.1). Now proceeding as is done for $\overline{U}_{2m}$, we conclude that $\overline{V}_e$ is nonempty. This contradiction with Theorem 2 completes the proof.

The regularity condition is essential for Theorem 3 to be valid. Indeed, consider the following

Example 6 The Z-code $X = \{a^{i+1} b a^i b : i = 0, 1, 2, \dots\} \subseteq \{a, b\}^*$ is not a regular language. It is not a code with bounded synchronization delay, although $A^* X^2 A^* \cap X = \emptyset$.

Concluding, from [8] or [1] we deduce the following statement.

Theorem 3 *Let* $X = \{x, y\} (|x| > |y|)$ *be a two-word language of* A^* *then* X *is not a* Z*-code if and only if one of the following assertions holds*

(i) x *or* y *is imprimitive;*

(ii) x *and* y *are conjugate;*

(iii) xy^n *is imprimitive for some positive integer* $n < \frac{|x|}{|y|} + 1$;

(iv) $x^2 y$ *is a square.*

Proof. Obviously, if one of (i)-(iv) holds, X is not a Z-code.

Conversely, suppose that X is not a Z-code (thus not a circular code, not a very pure code) and besides x and y are primitive and not conjugate. We show that (iii) or (iv) must occur.

Indeed, by [8] or [1], $x^* y \cup x y^*$ contains an imprimitive word $u = v^m, m \geq 2$:

– if $u = xy^n$ then $(n-1)|y|$ cannot excceed $|v| - 1$ otherwise by Fine and Wilf Theorem (see [9] or [5]) x and y are copower that contradicts the assumption. Thus $(n-1)|y| < |v|$, or $2(n-1)|y| < 2|v| \leq |x| + n|y| = |xy^n|$ i.e. $|n| < \frac{|x|}{|y|} + 1$;

– if $u = x^n y = v^m$ we can suppose $n \geq 2$. Further, if the inequality

$$m \geq \frac{n+1}{n-1}$$

holds, then $m(n-1)|x| \geq (n+1)|x| \geq n|x|+|y| = m|v|$. Therefore, $(n-1)|x| \geq |v|$, or, $n|x| \geq |x|+|v|$. Again by Fine and Wilf Theorem x, v and thus x, y are copower that contradicts the assumption. So, we always have $m < \frac{n+1}{n-1}$. Since $m, n \geq 2$ it follows $m = n = 2$.

Acknowledgements The authors acknowledge their great indebtedness to the referees for their valuable comments and suggestions.

References

[1] E. Barbin Le Rest and M. Barbin Le Rest, Sur la combinatoire des codes à deux mots, *Theoretical Computer Science* 41 (1985), 61-81.

[2] M.P. Beal, "Codages, automates locaux et entropie," Publications du LITP (Paris), No. 38 (1988).

[3] J. Berstel and D. Perrin, "Theory of Codes", Academic Press, New York, 1985.

[4] J. Devolder and E. Timmerman, "Codes for Bi-infinite Words", Publications du LIFL (Lille), No I.T. 16 (1990).

[5] N.J. Fine and H.S. Wilf, Uniqueness Theorems for Periodic Functions, *Proceedings of the American Mathematical Society* 16 (1965), 109-114.

[6] S. Golomb and B. Gordon, Codes with Bounded Synchronization Delay, *Information and Control* 8 (1965), 355-372.

[7] G. Lallement, "Semigroups and Combinatorial Applications," John Wiley, New York, 1979.

[8] A. Lentin and M.P. Shützenberger, A Combinatorial Problem in the Theory of Free Monoids, "Combinatorial Mathematics and Its Applications," Proceedings of the Conference held at the University of North Carolina (R.C. Bose and T.A. Dowlings, eds.), Chapell Hill, pp. 128-144.

[9] M. Lothaire, "Combinatorics on Words," Addison-Wesley, Reading, Massachusetts, 1983.

[10] Nguyen Huong Lam and Do Long Van, On a Class of Infinitary Codes, *Theoretical Informatics and Applications* 24 (1990), 441-458.

[11] A. Restivo, A Combinatorial Property of Codes Having Finite Synchronization Delay, *Theoretical Computer Science* 1 (1975), 95-101.

[12] A. Restivo, On a Question of McNaughton and Papert, *Information and Control* 25 (1974), 93-101.

Received May 9, 1992